United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
Number RDD-24-01

May 2024

Assessment of an Imputation Process Used in the 2017 Census of Agriculture

Tara Murphy
Habtamu Benecha
Denise A. Abreu
Darcy Miller

Assessment of an Imputation Process Used in the 2017 Census of Agriculture

Tara Murphy, Habtamu Benecha, Denise A. Abreu, Darcy Miller
National Agricultural Statistics Service, USDA,
1400 Independence Avenue, Washington DC 20250

Abstract

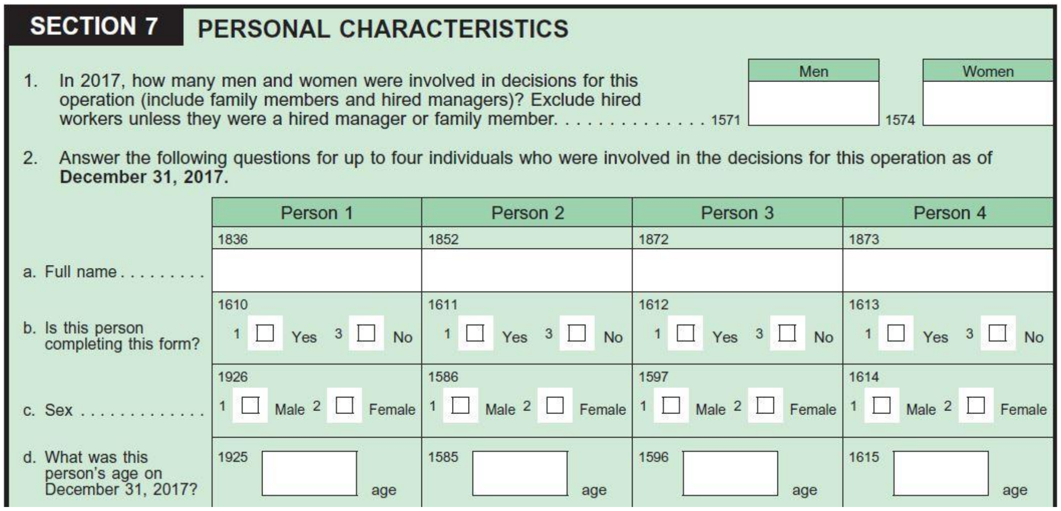
The United States Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) conducts the Census of Agriculture (COA) every five years using a list frame called the Census Mailing List (CML). The 2017 COA used a capture-recapture approach to adjust for undercoverage, nonresponse, and misclassification of farms. NASS used the June Area Survey (JAS), which is based on an area frame, as an independent survey in the capture-recapture approach. To apply this approach, a matched dataset is created by linking records from the CML and the JAS. The matched dataset is the foundation for modeling the probabilities of coverage, response, and correct classification of farms/non-farms in the COA. These probabilities are estimated through a series of weighted logistic regression models. Demographic characteristics are crucial predictors in these models. In 2017, NASS redesigned the demographics section of the COA questionnaire to allow reporting of up to four producers per farm. However, the JAS questionnaire gathered information on only one producer. Multivariate imputation was used to address this missing-data problem. This paper evaluates the impact of imputing the additional potential producers on the JAS on COA estimates.

Key words: Capture-recapture; Imputation; List frame; Area frame; Logistic regression

1. Introduction and Background

In United States (U.S.) agricultural surveys, Hispanic Americans, African Americans, and Amish farmers tend to be underrepresented (Escalante et al. 2006; Kraybill et al. 2013; Minkoff-Zern and Sloat 2017). In addition, female farmers, urban farmers, owners of small farms, and farmers with 10 or fewer years of experience (new/beginning farmers) tend to also have relatively low coverage on the National Agricultural Statistics Service's (NASS) Census of Agriculture (COA) list frame and thus suffer from underrepresentation. Since the release of the 2012 COA, there was increasing concern that the roles and contributions of women and new/beginning farmers needed to be better reflected in statistical statements about U.S. agriculture. Feedback to the United States Department of Agriculture (USDA) from the public, government sectors, and the Agricultural Statistics Advisory Committee identified the need to re-evaluate how NASS quantifies the contribution of women and new/beginning farmers in federally funded surveys.

In 2015, a panel of experts reviewed the COA to determine improvements that could be made to allow data users to better understand the role and effectiveness of USDA programs directed at women and new/beginning farmers. The panel recommended several updates to the COA questionnaire to achieve this goal. In response to one of the recommendations, NASS redesigned the demographics section of the 2017 COA questionnaire to allow up to four producers per farm ("Report of the Expert Panel" 2015) (Figure 1). By USDA definition, a farm is any place from which $1,000 or more of agricultural products were produced and sold, or normally would have been sold, during the COA year.



SECTION 7 PERSONAL CHARACTERISTICS

1. In 2017, how many men and women were involved in decisions for this operation (include family members and hired managers)? Exclude hired workers unless they were a hired manager or family member. . . . . 1571 Men 1574 Women

2. Answer the following questions for up to four individuals who were involved in the decisions for this operation as of **December 31, 2017.**

	Person 1	Person 2	Person 3	Person 4
a. Full name	1836	1852	1872	1873
b. Is this person completing this form?	1610 1 ☐ Yes 3 ☐ No	1611 1 ☐ Yes 3 ☐ No	1612 1 ☐ Yes 3 ☐ No	1613 1 ☐ Yes 3 ☐ No
c. Sex	1926 1 ☐ Male 2 ☐ Female	1586 1 ☐ Male 2 ☐ Female	1597 1 ☐ Male 2 ☐ Female	1614 1 ☐ Male 2 ☐ Female
d. What was this person's age on December 31, 2017?	1925 age	1585 age	1596 age	1615 age

Figure 1: 2017 COA demographics section snapshot.

The changes to the form collected a richer dataset for NASS to use in delivering products that better reflect the role of women and new/beginning farmers in agriculture. Downstream processing also had to be modified to absorb and utilize the richer dataset. For example, the proper use of demographic data in COA estimation and the weighting methodology to address the incompleteness of the COA list frame, called the Census Mailing List (CML).

To account for the incompleteness in the CML, estimation for the 2017 COA was conducted by applying a capture-recapture methodology that uses the COA and NASS's annual June Area Survey (JAS) as the two independent sources of data (Young et al.,2013, 2017; USDA National Agricultural Statistics Service, 2014). The JAS is a survey that uses an area-frame covering all land in the continental U.S. stratified by land use. The strata are further divided into substrata by grouping areas that are agriculturally similar. Within each substratum, the land is divided into primary sampling units (PSUs). A sample of PSUs is selected, and smaller, similar-sized segments of land are sampled from the selected PSUs to be fully enumerated (white outlined area in Figure 2). Each segment consists of tracts of land representing unique land operating arrangements. Before the data collection period, all tracts of land within selected segments (black outlined areas with letter labeling in Figure 2) are screened and classified as agricultural or non-agricultural (Lamas et. Al, 2010). A JAS questionnaire is completed only on those tracts identified as agricultural and their farm status is determined per the farm definition.

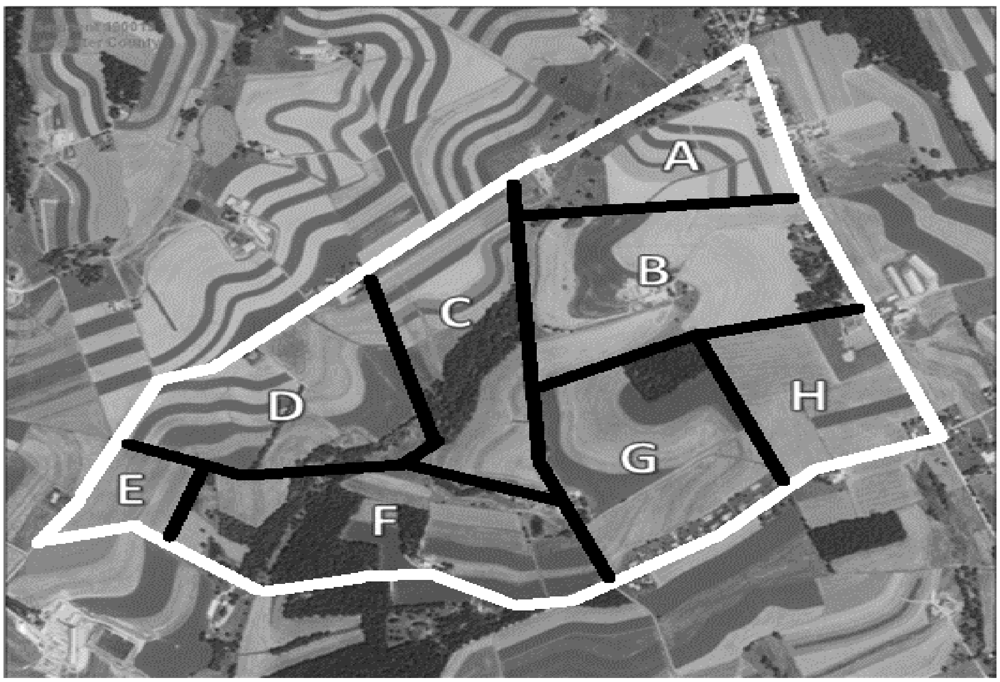
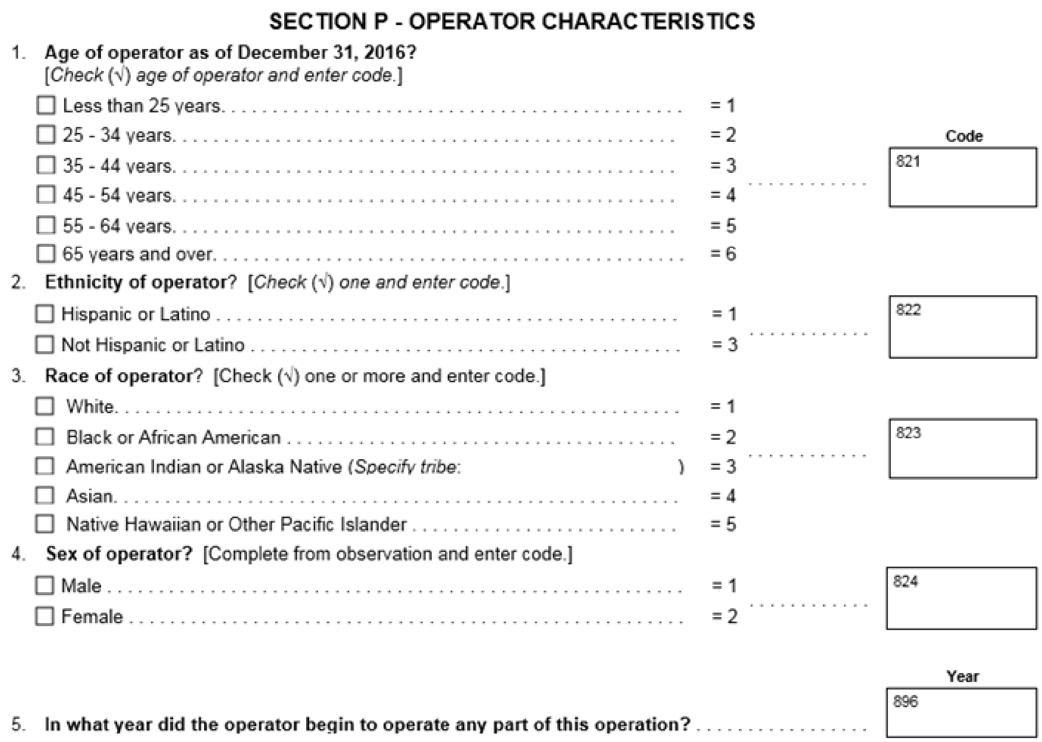


Figure 2: An example JAS segment with tracts A-H.

To implement the capture-recapture approach, the CML and the JAS are linked using probabilistic record linkage. The matched dataset consisting of JAS and CML records is then used in logistic regression models for estimating the different adjustment weights for all responding COA records. Demographic characteristics are crucial covariates in the logistic regression models used for producing COA adjustment weights.

While the 2017 COA questionnaire allowed reporting demographic characteristics for up to four producers per farm, the JAS questionnaire collected information on only one operator (Figure 3), the person who makes most of the day-to-day decisions (one of the decision-making questions on the COA). For purposes of simplicity, the JAS operator will henceforth be referred to as a "producer." Ideally, the demographic information on the 2017 JAS would have been collected in the same manner as the 2017 COA for model-based estimation of the different COA adjustment weights. When the CML and the JAS matched dataset was created, the demographic variables associated with producers 2, 3, and 4 were missing for the JAS records. JAS records are a crucial element for modeling coverage of the CML. Because COA publications include demographic estimates at the county level, it is essential for the demographic variables to be included in the COA model.



SECTION P - OPERATOR CHARACTERISTICS

1. **Age of operator as of December 31, 2016?**
[*Check (√) age of operator and enter code.*]
☐ Less than 25 years. . . . . = 1
☐ 25 - 34 years. . . . . = 2
☐ 35 - 44 years. . . . . = 3
☐ 45 - 54 years. . . . . = 4
☐ 55 - 64 years. . . . . = 5
☐ 65 years and over. . . . . = 6
Code: 821

2. **Ethnicity of operator**? [*Check (√) one and enter code.*]
☐ Hispanic or Latino . . . . = 1
☐ Not Hispanic or Latino . . . . = 3
822

3. **Race of operator**? [Check (√) one or more and enter code.]
☐ White. . . . . = 1
☐ Black or African American . . . . = 2
☐ American Indian or Alaska Native (*Specify tribe*:) = 3
☐ Asian. . . . . = 4
☐ Native Hawaiian or Other Pacific Islander . . . . = 5
823

4. **Sex of operator?** [Complete from observation and enter code.]
☐ Male . . . . = 1
☐ Female . . . . = 2
824

5. **In what year did the operator begin to operate any part of this operation?** . . . .
Year: 896

Figure 3: 2017 JAS demographics section snapshot.

Previous studies (Ridolfo 2015; Pick et al. 2016; Ridolfo et al. 2016) have shown the type of producer that is typically reported first and provide insight on the benefits of allowing additional producers to be reported. By imputing the missing producers for the JAS records, the objective was to increase the data available for COA modeling and allow the JAS data to be more reflective of the population, which was the goal of changing the COA questionnaire.

This paper highlights the impact of the imputation conducted on the JAS records to expand the number of potential producers on the JAS to four on the 2017 COA estimates for selected demographic groups, which include female, young (age less than or equal to 35), new/beginning, and non-white producers. First, the COA estimation methods are briefly discussed in Section 2. In Section 3, the demographic imputation methods are outlined. The approaches used to evaluate the impacts of imputation and the results are then presented in Sections 4 and 5, respectively. We conclude with a discussion in Section 6.

2. Estimation from the Census

The COA is a list-based endeavor. In preparation for the quinquennial COA, the list of farm operations from the previous COA is updated and new and potential farms are added based on information from other sources. Shortly before the COA, this list frame is “frozen” (no additional farms or potential farms are added or removed) and becomes the CML. During the COA, a questionnaire is sent to each operation on the CML. The CML contains both agricultural operations that are in the target population (farms) and agricultural operations that are not in the target population (non-farms). The CML is incomplete; not all farms are on the list. To account for farming operations not on the CML, NASS uses the JAS. The COA uses capture-recapture methodology to adjust COA respondents for various sources of error. A matched dataset consisting of all matches of a CML record to a JAS tract is formed. The matching is performed using probabilistic record linkage. For the 2017 COA, this dataset was the foundation for modeling the probabilities of coverage, response, and misclassification of farms in the COA. These probabilities were estimated through a series of weighted logistic regression models. There are two types of misclassifications in the COA; a farm can be counted as a non-farm (undercounting), or a non-farm can be counted as a farm (overcounting). Overcounting and undercounting adjustment weights were estimated from two separate logistic regression models for the 2017 COA. COA estimates were obtained by applying the four adjustment weights (i.e., undercoverage, nonresponse, undercounting, and overcounting) to responding COA records. Model-based estimates from the COA are then calibrated before the results are published (Sartore et. al, 2019).

Demographic characteristics are crucial covariates for the COA estimation models. The 2017 COA collected demographic data from up to four producers for each farm, but the JAS questionnaire allowed for only one producer. This mismatch in the number of producers per farm collected from the COA versus the JAS made it challenging to define demographic covariates for each farm. For this reason, NASS imputed demographic information for up to three additional producers on the JAS form using donors from the COA administered in the same year as discussed in the next section.

3. Imputation Approach

Hot deck imputation often describes a general class of imputation methods that utilize the current survey data observations (the 'hot' data) to impute data. Often it is implemented as a process where groups of 'like' records are formed and a respondent value is drawn from the same group as the recipient to provide an imputed value for the recipient.

In this implementation of the hot deck method, no donors were available in the current JAS survey to use to impute demographic items for more than one producer for other JAS records. Thus, 2017 COA demographic data were added to the pool of donors for imputing producers 2, 3, and 4 on the JAS records. Groups were formed based on the values of the producer collected on the JAS form and the producer listed in the first column of the COA questionnaire. Demographic variables used to form similar groups included age, race, and sex of the first producer listed. An entire COA record was drawn from the group to impute producers 2, 3, and 4 on the JAS. Using the entire COA record as a donor, the distributions of the number of producers and joint demographics of producers were maintained. The distribution of the number of producers was preserved since records drawn would have zeros as placeholders for variables collecting information on producers beyond the number of producers on the farm. For example, the demographic values on the COA record drawn could all be 0s, meaning that the COA record only had one producer, ensuring that the distribution of single producer farms was still preserved in the JAS. Demographic values drawn for producers 2, 3, and 4 could all be zero except for values corresponding to the second producer, preserving the distribution of two producer farms, and similarly for three and four producer farms. Any items requiring imputation in the data for the one producer that was collected on the JAS form was imputed using other JAS records where all of the items were reported before imputing data for potential additional producers.

Imputation was implemented using the PROC SURVEYIMPUTE procedure available in the SAS software (SAS/STAT 14.1 User's Guide, 2015). A donor is selected for a recipient unit, and the observed values of the donor are imputed for the missing items of the recipient. Available donor selection techniques include simple random selection with or without replacement, probability proportional to weights selection (Rao and Shao, 1992), and approximate Bayesian bootstrap selection (Rubin and Schenker, 1986). For the JAS imputation process, simple random selection with replacement was used. As expected, this method yielded a joint distribution of demographic data similar to the joint distribution of the demographic data on the COA.

4. Evaluation of the impact of imputation

For the 2017 COA, an imputed JAS dataset was used with COA records for producing model-based estimates. This paper is focused on evaluating the impacts of JAS imputation on COA estimates, with a particular focus on its effect on the demographic characteristics of producers. Based on studies conducted to redesign the 2017 COA demographics section, there was an expectation of capturing more female and young producers (Ridolfo et al. 2016).

4.1. Case study: Preliminary evaluation

To assess impacts of the imputation on COA estimates, Murphy et al. (2019) used the 2017 production COA data and made comparisons between the 2017 COA model estimates with and without JAS imputation for the demographic variables associated with producers 2, 3, and 4 on the JAS. The characteristics evaluated were age, sex, race, and ethnicity. Preliminary results showed that both the number of farms with young producers and the number of farms with at least one female producer were generally larger when imputation efforts were applied. While this was the general result, it did not hold for a small number of domains (Murphy et al. 2019).

4.2. Simulation study

Following the case study analysis of Murphy et al. (2019) based on the 2017 COA production data, formal attempts were made to study the impacts of JAS demographic imputation on demographic estimates from the COA by using simulated JAS demographic data.

A total of 500 JAS datasets were simulated in such a way that each dataset preserved demographic characteristics in the JAS. For the study, these 500 simulated replicates were used as control datasets representing COA estimation without JAS imputation for producers 2, 3, and 4. Then, each of those control datasets received a treatment (imputed demographic characteristics for producers 2, 3, and 4 using the imputation method used during the 2017 COA cycle, as outlined in Section 3), yielding 500 treatment datasets. Each of the 500 JAS datasets initially simulated (i.e., without additional demographic imputation) and the 500 treated JAS datasets (i.e., with additional demographic imputation) were linked to the CML records to create a matched dataset for producing estimates by applying the same capture-recapture procedures as the 2017 COA estimation, including stepwise variable selection. Comparisons were then made between model estimates based on the 500 simulated JAS datasets that were not imputed for demographic characteristics and the 500 simulated datasets that were imputed (Figure 4). While published COA estimates are obtained by calibrating model-based estimates, calibration is ignored in our comparisons to avoid the effects of possible adjustments and noise in the estimates.

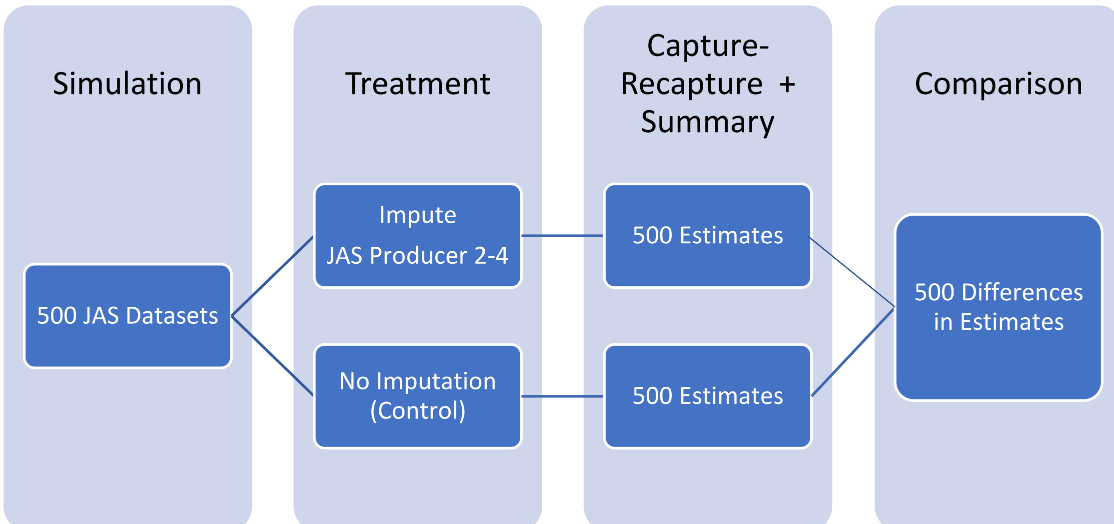


Figure 4: Diagram describing simulation study.

5. Results

While estimates were produced for all demographic groups, this paper focuses on comparisons of estimates with and without imputation for the numbers of farms with at least one (1) female producer, (2) young producer, (3) new/beginning producer, and (4) non-white producer, as well as the total numbers of (5) female producers, (6) young producers, (7) new/beginning producers, and (8) non-white producers.

Analysis of record-level model estimated weights showed that some responding COA records had very high weights for some of the simulations. These weights resulted in unrealistically high estimates, particularly at the county level. In the absence of calibration as in the official estimates from the COA, weights exceeding 100 were removed before summaries were produced from the models based on both imputed and unimputed datasets. The cutoff value of 100 was chosen based on analysis of the distribution of weights from each simulation. The percentages of records removed from the unimputed datasets range from 0.0002% to 0.0023%. For the imputed datasets, the percentages of records removed from the analysis range from 0.029% to 0.061%.

After the model-based estimates were produced from each of the remaining replicates, percent relative differences between estimates from the imputed and unimputed datasets were computed for the eight demographic characteristics listed above.

$$\text{Percent relative difference} = 100 \times \frac{\text{imputed} - \text{not imputed}}{\text{not imputed}}$$

To gauge geographic differences, states in the continental US were grouped into seven regions based on similarity of agriculture using feedback from NASS's subject matter experts (Figure 5). The regions are defined as follows:

Region 1: Connecticut, Maine, Massachusetts, Michigan, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Wisconsin.

Region 2: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, South Carolina.

Region 3: Delaware, Kentucky, Maryland, North Carolina, Tennessee, Virginia, West Virginia.

Region 4: Illinois, Indiana, Iowa, Kansas, Minnesota, Missouri, Montana, Nebraska, North Dakota, Ohio, Oklahoma, South Dakota.

Region 5: Arizona, Colorado, Idaho, Nevada, New Mexico, Utah, Wyoming.

Region 6: California, Oregon, Washington.

Region 7: Texas.

National (i.e., U.S.) estimates were obtained from these regions. In addition, to assess how imputation affects estimates of demographic groups of interest versus estimates of contrasting demographic characteristics (e.g., female vs. male), percent relative differences between estimates from the imputed and unimputed datasets were calculated for each of the 500 pairs of replicates.

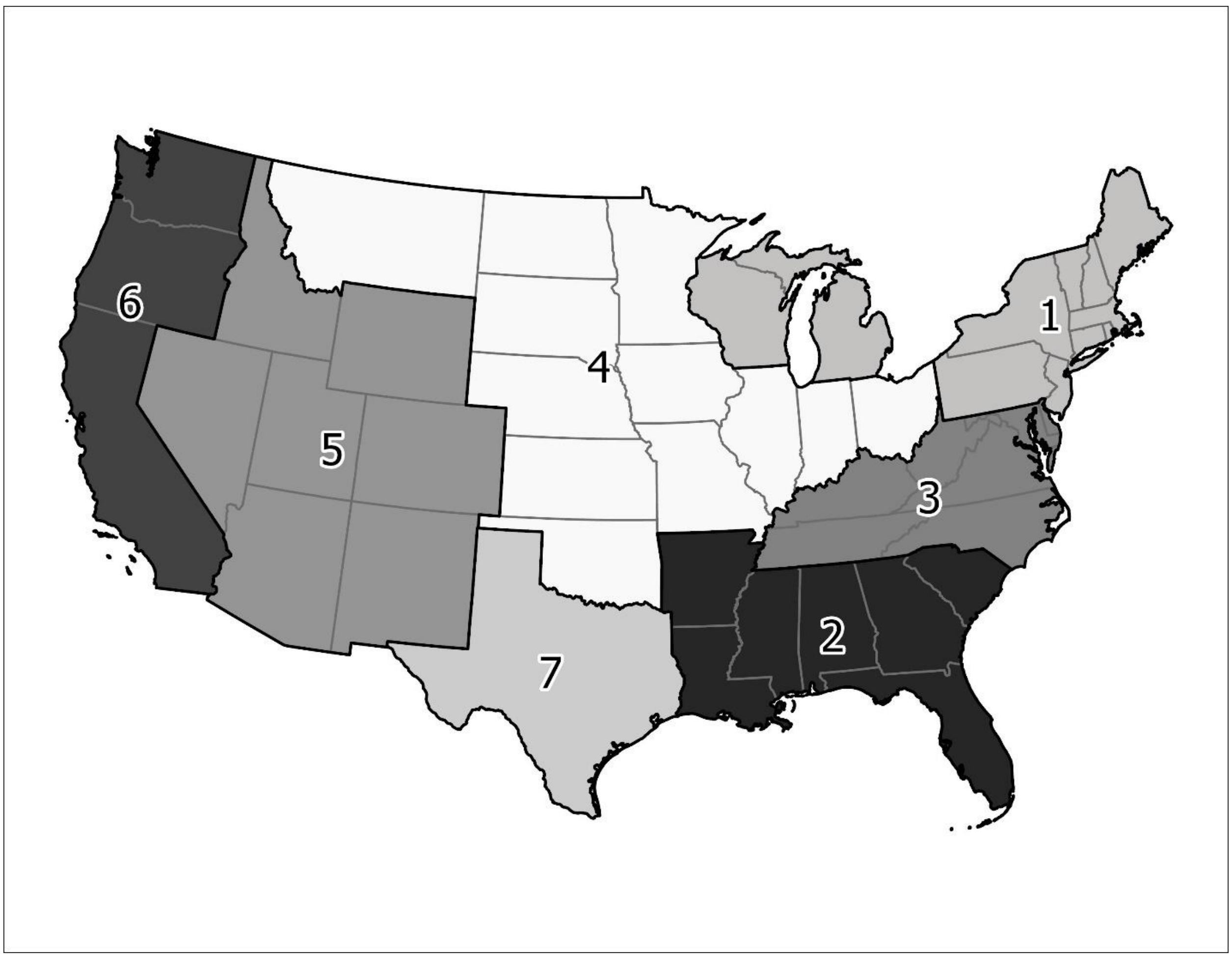


Figure 5: States included in seven agricultural regions studied.

As shown in the map in Figure 6, the smallest regional percent relative difference between estimates from imputed and unimputed datasets for the number of farms with at least one female producer was 18.4%. In the map, regions with a light fill have the smallest percent differences, and as the color gets darker the percent differences get larger. For the number of farms with at least one female producer, the largest percent difference was 32.8% for Region 5. At the U.S. level, percent relative difference of number of farms with at least one female producer was 24.1%. Thus, imputation of demographic characteristics in the JAS provided higher estimates compared to analysis of the data without imputation of the JAS data.

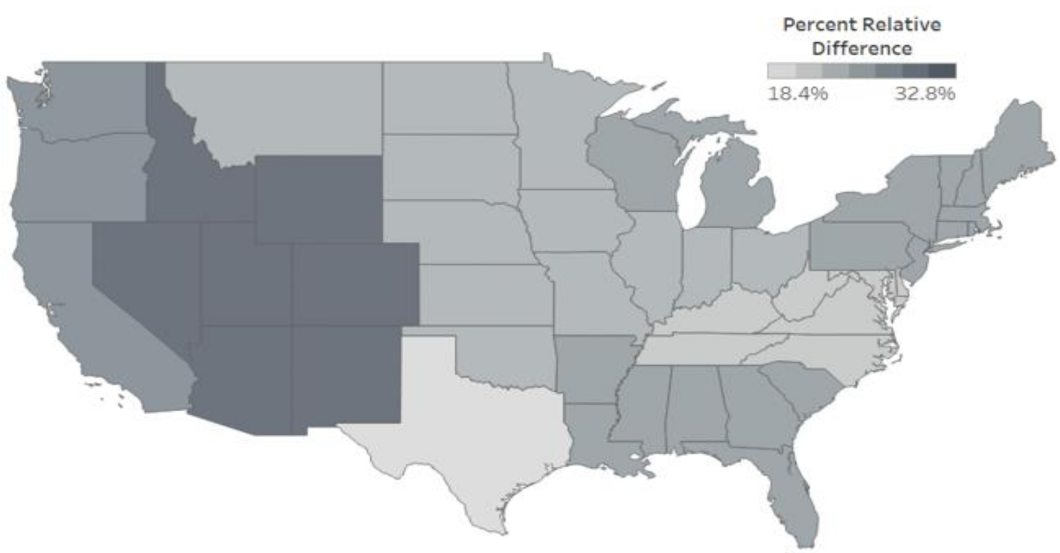


Figure 6: Percent relative difference between estimates from imputed and unimputed datasets for the number of farms with at least one female producer.

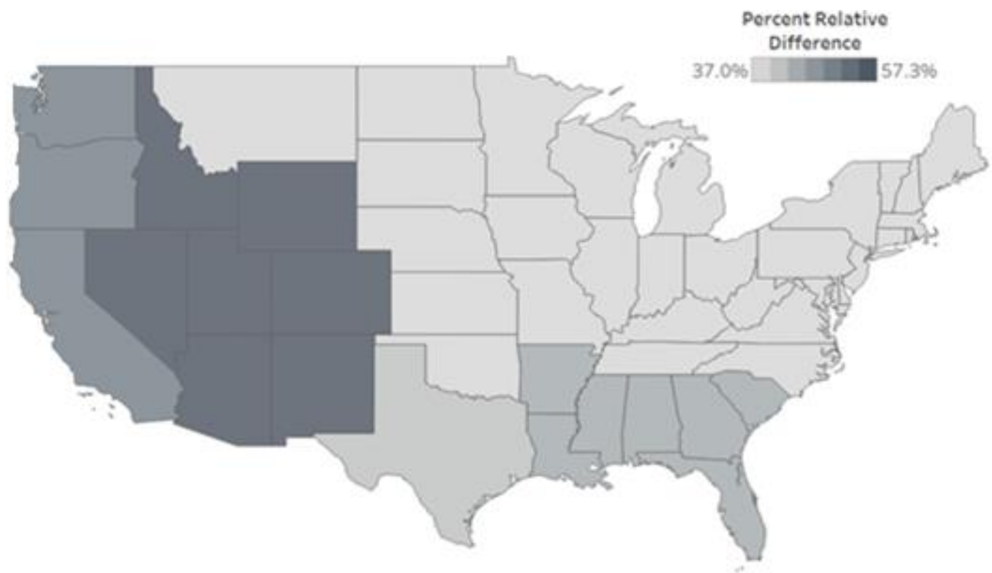


Figure 7: Percent relative difference between estimates from imputed and unimputed datasets for the number of farms with at least one young producer.

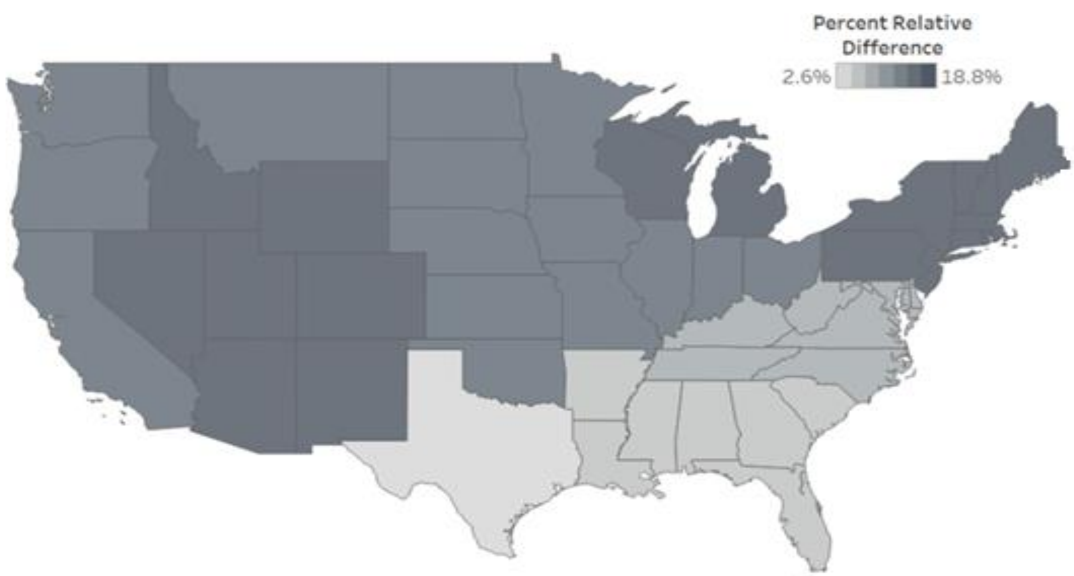


Figure 8: Percent relative difference between estimates from imputed and unimputed datasets for the number of farms with at least one new/beginning producer.

The regional percent relative differences for the number of farms estimates with at least one young producer ranged from 37% to 57.3% (Figure 7). The corresponding percent differences for the number of farms with at least one new/beginning producer ranged from 2.6% to 18.8% (Figure 8). The percent relative differences at the U.S. level were 41.0% and 11.5%, respectively, for the number of farms with at least one young producer and those with at least one new/beginning producer. For the number of farms with at least one non-white producer, the regional relative difference between estimates from the imputed and unimputed datasets range from 0.3% to 38.3% (Figure 9), and the U.S. level percent relative difference was 9.7%.

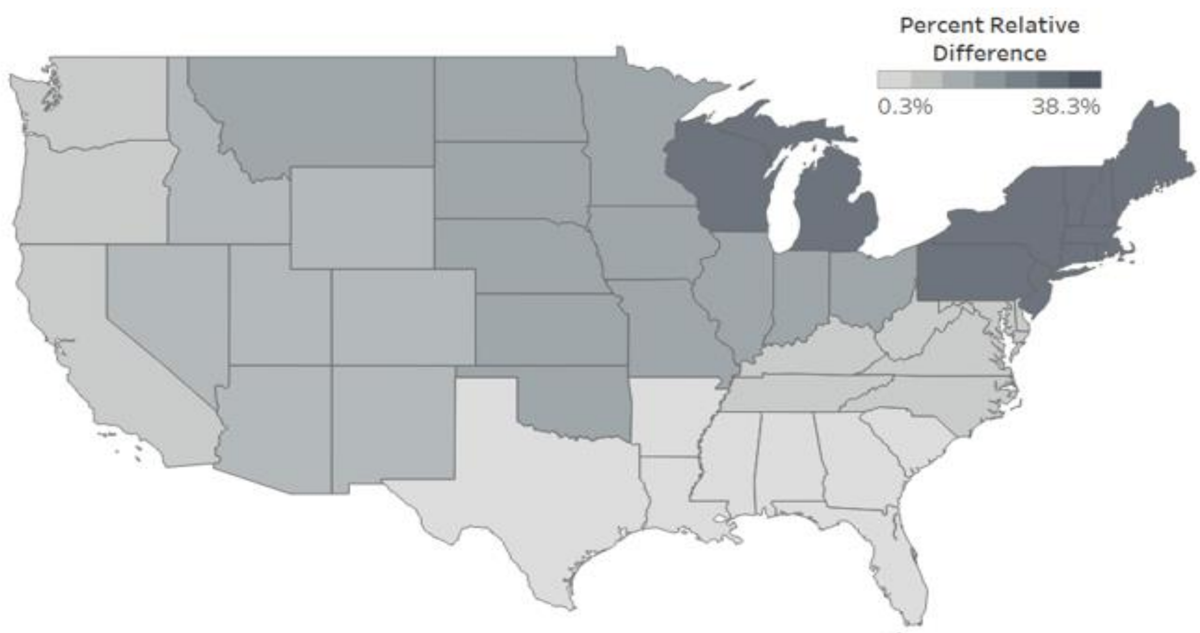


Figure 9: Percent relative difference between estimates from imputed and unimputed datasets for the number of farms with at least one non-white producer.

Similarly, U.S. level percent relative differences between estimates from imputed and unimputed datasets were 26.2%, 39.9%,17.4%, and 16.2%, respectively, for the total numbers of young, female, new/beginning, and non-white producers. As the maps in Figures A1-A4 in the Appendix also show, the regional percent relative differences for the total numbers of young, female, new/beginning, and non-white producers were all at least 6.5%.

5.1. Overall findings

JAS demographic imputation provided COA estimates that are higher, compared to estimates obtained without imputation, although there are differences among states and regions on the impact of imputation on the estimates. Our findings from the simulation analysis were consistent with the 2019 case study and the general expectation that imputation would reflect more representation of some demographic groups in the COA.

Results from the simulation study show that JAS imputation not only increases the number of farms and total number of producers with the demographic characteristics discussed above, but it also increases estimates for almost all COA categories. For example, the average total number of farms from imputed datasets are greater than those from unimputed datasets. As the boxplots in Figures A5-A8 in the Appendix also show, the numbers of farms with male producers, older producers (older than 35), more established producers (operating more than 10 years), and white producers were higher from the imputed datasets compared to the unimputed datasets; however, the impact of imputation is more substantial on the estimates for female, young, new/beginning, and non-white producers.

6. Discussion

Following recommendations from a panel of experts who reviewed the COA, NASS redesigned the demographics section of the 2017 COA questionnaire to allow reporting for up to four producers per farm ("Report of the Expert Panel" 2015). Thus, up to four producers were recorded the 2017 COA for each farm record. However, the JAS, which is used to adjust for undercoverage in the CML, collected demographic data on only one producer.

The mismatch in the number of producers per farm collected from the COA and the JAS made it challenging to define demographic covariates for each farm to be used in modeling COA estimates. As a result, NASS imputed demographic information for up to three additional producers on the JAS form using donors from the COA administered in the same year.

Preliminary analysis based on production 2017 COA data showed that imputation of JAS demographic characteristics generally increases estimates of the numbers of farms operated by young and female producers (Murphy et al. 2019). This research was expanded using simulation studies. A total of 500 JAS data sets without imputation and 500 datasets with imputation were generated, and each JAS dataset was linked to the CML. Estimates were produced by applying the same estimation procedures used for the 2017 COA. Comparisons of the estimates from the simulation studies show that JAS imputation increases the number of farms operated by producers of different demographic characteristics relative to results obtained without JAS imputation. While estimates obtained from imputed datasets are generally higher than those from unimputed datasets, the increase in estimates is more substantial in the number of farms with at least one young producer, female producer, new/beginning producer, and non-white producer. The results indicate that allowing for more producers to be reported on the JAS will lead to finding more farms with at least one young producer and at least one female producer, among other minority groups. For the 2022 COA, which results were just recently released, the JAS questionnaire was redesigned to allow up to four producers to be reported, matching the constitution of the 2022 COA questionnaire.

The general national and international population relies on U.S. agriculture to help feed and clothe the world. Agricultural data are a part of U.S. national security, consumer protections, trade, conservation and environmental quality, education, supply chain operations and recreational program planning. Findings in this study are important to understanding the current state of agriculture. The USDA, U.S. Congress, and other stakeholders need to understand the current state of U.S. agricultural producers to plan for the future of U.S. agriculture.

Acknowledgements

The findings and conclusions in this report are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy.

The authors would like to acknowledge Virginia Harris for sharing subject matter expertise on producer demographics, Luca Sartore for assistance with calibration, Sarah Goodale for providing knowledge on the processes used for 2017 COA estimate production, and Arthur Rosales for creating the map of regions evaluated in the simulation study.

7. References

Abreu, D. A., J. S. McCarthy, and L. A. Colburn (2010). Impact of the Screening Procedures of the June Area Survey on the Number of Farms Estimates. Research and

Development Division. RDD Research Report #RDD-10-03. Washington, DC: USDA, National Agricultural Statistics Service.

Benecha, H. K., Abreu, D. A., Jennings, R., & Young, L. J. Data reconciliation and estimation in an agricultural survey. Statistical Journal of the IAOS, (Preprint), 1-12.

Lamas, A. C., Abreu, D. A., Arroway, P., Lopiano, K. K., & Young, L. J. (2010). Modeling misclassification in the June Area Survey. In Proceedings of the 2010 Joint Statistical Meetings, ASA Section on Survey Research Methods (pp. 2480-2488).

Miller, Darcy. (2018). “Dancing with a New Partner: Imputing New Demographic Questions on the Census of Agriculture Using Commercial-off-the-Shelf (COTS) Software,” 2018 Joint Statistical Meetings Proceedings.

Murphy, Tara, Habtamu Benecha, Denise A. Abreu, Darcy Miller. (2019). “Assessment of an Imputation Process Used in the 2017 Census of Agriculture,” 2019 Joint Statistical Meetings Proceedings.

N. K. Rao, J & Shao, J. (1992). Jackknife Variance Estimation with Survey Data Under Hot Deck Imputation. Biometrika. 79. 10.2307/2337236.

Pick, Kenneth, Kathy Ott, Heather Ridolfo, Jaki McCarthy, Rachel Sloan, Jeremy Beach, and Shirley Samson. (2016). 2015 Census of Agriculture Content Test: Results from Round 2 Cognitive Testing. Washington, DC: National Agricultural Statistics Service.

Report of the Expert Panel on Statistics on Women and Beginning Farmers in the USDA Census of Agriculture. 2015. Unpublished report.

Ridolfo, Heather. (2015). Census of Agriculture Section 6 Personal Characteristics: Results from Cognitive Testing. Fairfax, VA: National Agricultural Statistics Service.

Ridolfo, Heather, Emilola Abayomi and Virginia Harris. (2016). “Challenges to Developing New Survey Questions: When Cultural Norms Run Counter to Survey Questions.” Paper presented at the International Conference on Questionnaire Design, Development, Evaluation, and Testing (QDET2), Miami, FL.

Rubin, Donald and Nathaniel Schenker (1986). “Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse.” Journal of the American Statistical Association. Vol. 81, No. 394, Pp. 366-374.

Sartore L, Toppin K, Young LJ, Spiegelman C. Developing integer calibration weights for Census of Agriculture. Journal of Agricultural, Biological and Environmental Statistics. 2019; 24(1), 26-48.

SAS Institute, Inc. (2015). SAS/STAT 14.1 Users Guide. PROC SURVEYIMPUTE Statement :: SAS/STAT(R) 14.1 User's Guide

Sloan, Rachel, Heather Ridolfo, Zulma Riberas, Jaki McCarthy, Kathy Ott, and Shirley Samson. (2015). 2015 Census of Agriculture Content Test – Results from Round 1 Cognitive Testing. Fairfax, VA: National Agricultural Statistics Service.

USDA National Agricultural Statistics Service (NASS). U.S. Census of Agriculture: United States Summary and State Data, Vol. 1: Geographic Area Series, Part 51. Report AC-12-A-51., U.S.D.A.: Washington DC. 2014. Online: https://agcensus.library.cornell.edu/wp-content/uploads/usv1.pdf. Accessed 11 March 2024.

Young, Linda J., Lamas, Andrea C. and Abreu, Denise A. (2017). “2012 Census of Agriculture: A Capture-Recapture Analysis.” Journal of Agricultural, Biological, and Environmental Statistics. 22:523-539.

Young, L.J., Lamas, A.C., Abreu, D.A., Wang, S., Adrian, D. "Statistical methodology for the 2012 U.S. Census of Agriculture." In the Proceeding 59th ISI World Statistics Congress. 2013; 1063-1068.

8. Appendix

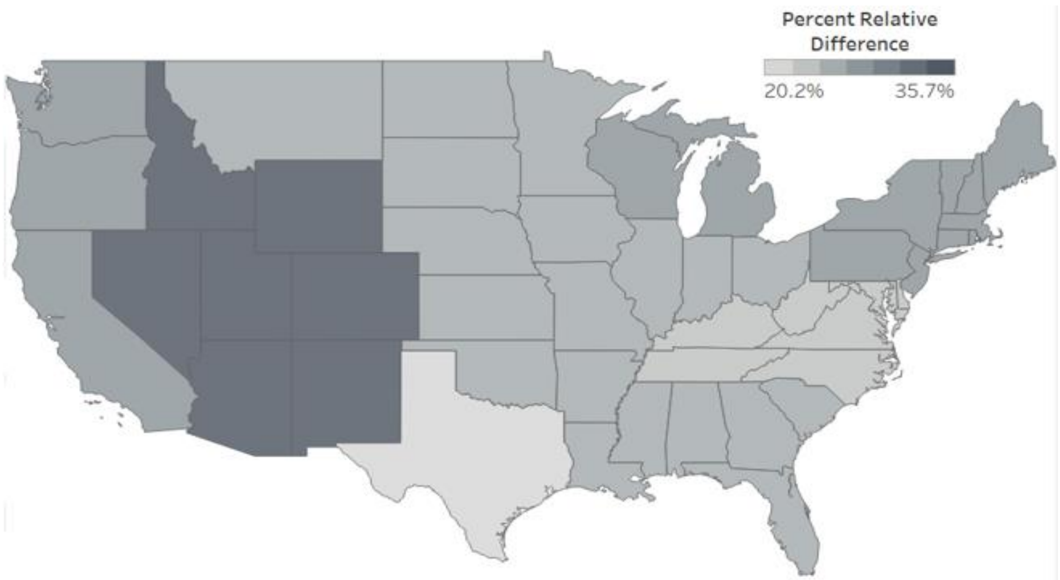


Figure A1: Percent relative difference between estimates from imputed and unimputed datasets for the number of female producers.

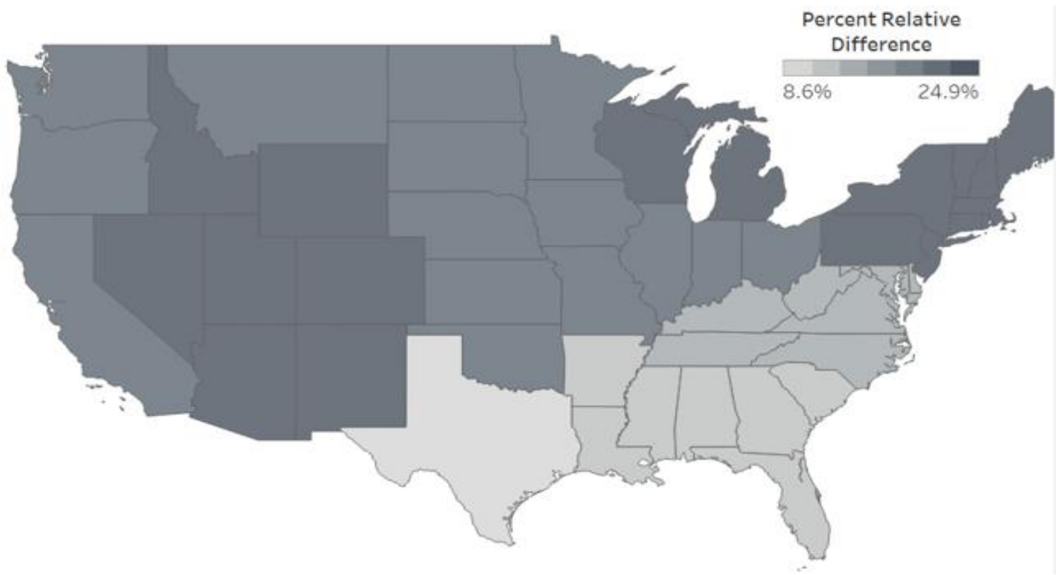


Figure A2: Percent relative difference between estimates from imputed and unimputed datasets for the number of new/beginning producers.

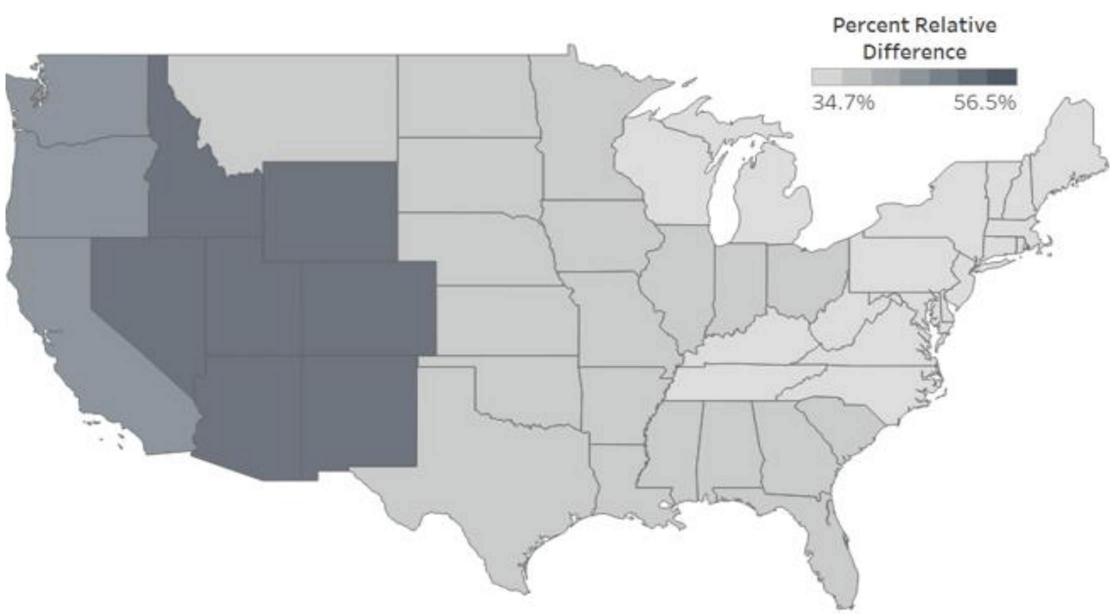


Figure A3: Percent relative difference between estimates from imputed and unimputed datasets for the number of young producers.

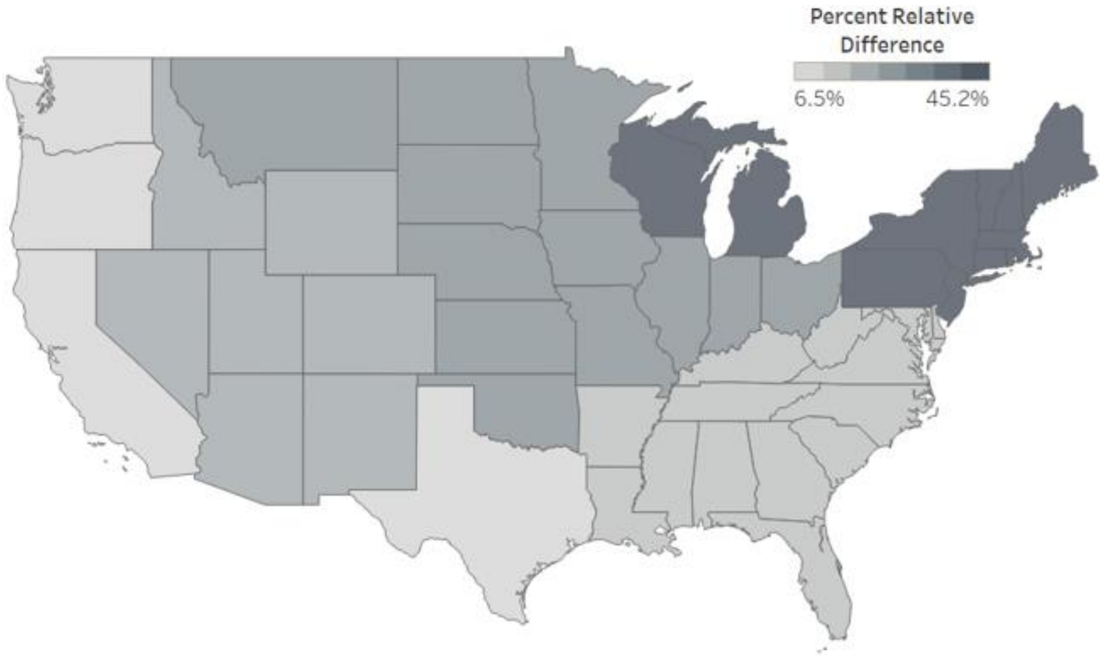


Figure A4: Percent relative difference between estimates from imputed and unimputed datasets for the number of non-white producers.

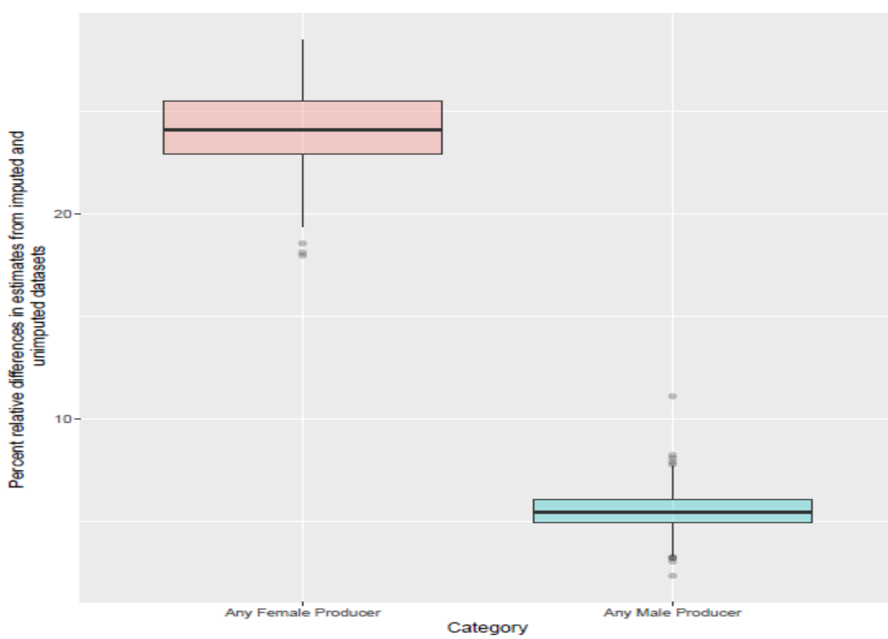


Figure A5: Percent relative differences in estimates of the numbers of farms operated by at least one female producer (at least one male producer) between models using imputed and unimputed datasets based on 500 simulations.

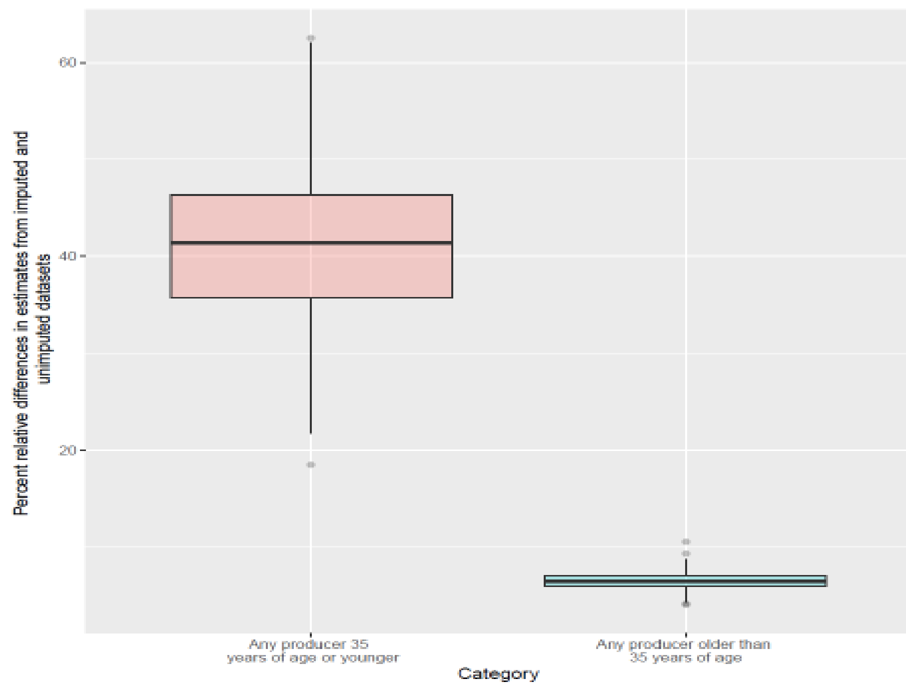


Figure A6: Percent relative differences in estimates of the numbers of farms operated by at least one producer aged 35 or younger (older than 35) between models using imputed and unimputed datasets based on 500 simulations.

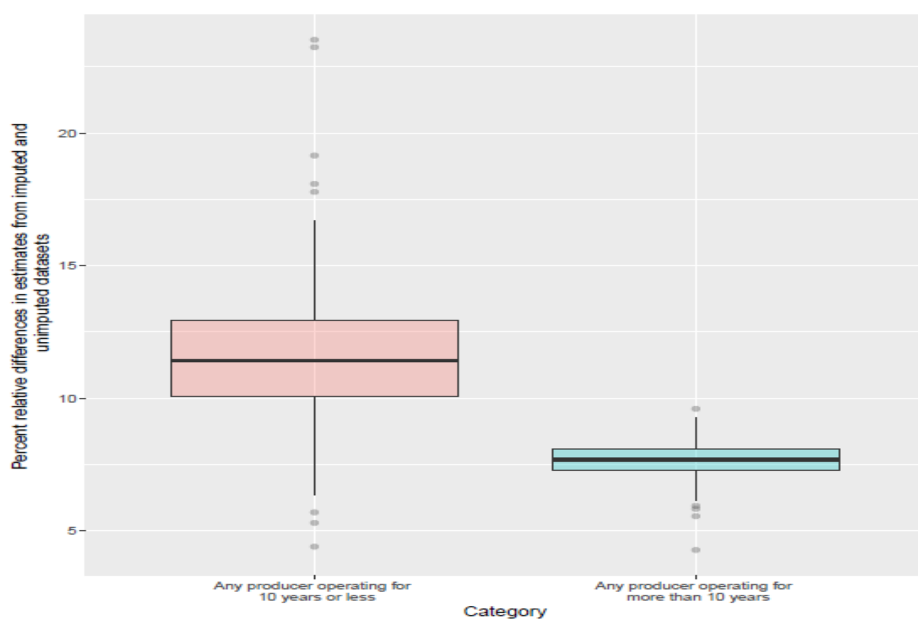


Figure A7: Percent relative differences in estimates of the numbers of farms operated by at least one producer operating for 10 years or less (more than 10 years) between models using imputed and unimputed datasets based on 500 simulations.

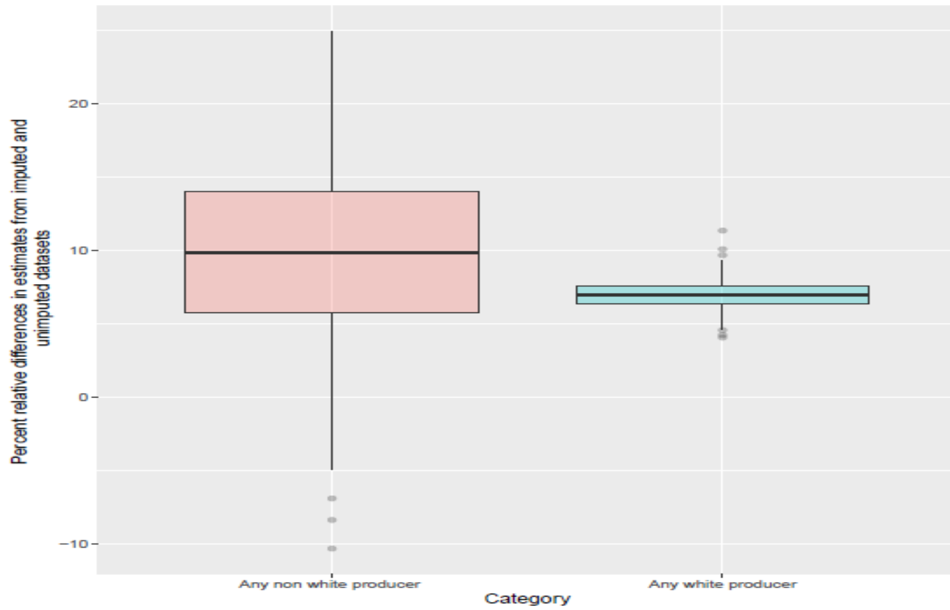


Figure A8: Percent relative differences in estimates of the numbers of farms operated by at least one non-white producer (white producer) between models using imputed and unimputed datasets based on 500 simulations.